1   **Single cell derived mRNA signals across human kidney tumors**

2

3   Matthew D Young[†*1], Thomas J Mitchell[†1,2,3], Lars Custers[†4,5], Thanasis Margaritis[4], Francisco

4   Morales[4,5], Kwasi Kwakwa[1], Eleonora Khabirova[1], Gerda Kildisiute[1], Thomas RW Oliver[1,2],

5   Ronald R. de Krijger[4], Marry M. van den Heuvel-Eibrink[4], Federico Comitani[7,8,9], Alice Piapi[10],

6   Eva Bugallo-Blanco[10], Christine Thevanesan[10], Christina Burke[10], Elena Prigmore[1], Kirsty

7   Ambridge[1], Kenny Roberts[1], Felipe A Vieira Braga[11], Tim HH Coorens[1], Ignacio Del Valle[10],

8   Anna Wilbrey-Clark[1], Lira Mamanova[1], Grant D Stewart[2,3], Vincent J Gnanapragasam[2,3,13],

9   Dyanne Rampling[12], Neil Sebire[10,12], Nicholas Coleman[2,6], Liz Hook[2,6], Anne Warren[2],

10   Muzlifah Haniffa[1,14,15], Marcel Kool[4,16,17], Stefan M Pfister[16,17,18], John C Achermann[10],

11   Xiaoling He[19,20], Roger A Barker[19,20], Adam Shlien[7,8,9], Omer A Bayraktar[1], Sarah Teichmann[1],

12   Frank C. Holstege[4], Kerstin B Meyer[1], Jarno Drost[4*], Karin Straathof[10,12*], Sam Behjati[1,2,21*]

13

14   **Affiliations**

15   [1]Wellcome Sanger Institute, Hinxton CB10 1SA, UK.

16   [2]Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK.

17   [3]Department of Surgery, University of Cambridge, Cambridge, CB2 0QQ, UK.

18   [4]Princess Máxima Center for Pediatric Oncology, 3584 CS Utrecht, The Netherlands.

19   [5]Oncode institute, 3584 CS Utrecht, The Netherlands.

20   [6]Department of Pathology, University of Cambridge, Cambridge, CB2 1QP, UK.

21   [7]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada.

22   [8]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario,

23   Canada.

24   [9]Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario,

25   Canada.

26   [10]UCL Great Ormond Street Hospital Institute of Child Health, London WC1N 1EH, UK.

27   [11]Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.

28   [12]Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK.

29   [13]Cambridge Urology Translational Research and Clinical Trials office, Cambridge Biomedical Campus

30   Cambridge CB2 0QQ University of Cambridge

31   [14]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals

32   NHS Foundation Trust, Newcastle upon Tyne, NE1 4LP, UK.

33   [15]Intitute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, NE1 4HH, UK.

34   [16]Hopp Children´s Cancer Center Heidelberg (KiTZ), 69120 Heidelberg Germany.

35   [17]German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Division of

36   Pediatric Neurooncology, 69120 Heidelberg, Germany.

37   [18]Heidelberg University Hospital, Department of Pediatric Hematology and Oncology, 69120

1   Heidelberg, Germany.

2   [19]MRC-WT Cambridge Stem Cell Institute, University of Cambridge, CB2 0QQ, Cambridge, UK.

3   [20]Department of Clinical Neuroscience, University of Cambridge, Cambridge, CB2 0QQ, UK.

4   [21]Department of Paediatrics, University of Cambridge, Cambridge, CB2 0QQ, UK.

5

6   [†]Joint first authors.

7   [*]Corresponding authors: sb31@sanger.ac.uk; my4@sanger.ac.uk; k.straathof@ucl.ac.uk;

8   j.drost@prinsesmaximacentrum.nl

9

10

11

Abstract

The cellular transcriptome may provide clues into the differentiation state and origin of human cancer, as tumor cells may retain patterns of gene expression similar to the cell they derive from. Here, we studied the differentiation state and cellular origin of human kidney tumors, by assessing mRNA signals in 1,300 childhood and adult renal tumors, spanning seven different tumor types. Using single cell mRNA reference maps of normal tissues generated by the *Human Cell Atlas* project, we measured the abundance of reference "cellular signals" in each tumor. Quantifying global differentiation states, we found that, irrespective of tumor type, childhood tumors exhibited fetal cellular signals, thus replacing the long-held presumption of "fetalness" with a precise, quantitative readout of immaturity. By contrast, in adult cancers our assessment refuted the suggestion of dedifferentiation towards a fetal state in the overwhelming majority of cases, with the exception of lethal variants of clear cell renal cell carcinoma. Examining the specific cellular phenotype of each tumor type revealed an intimate connection between the different mesenchymal populations of the developing kidney and childhood renal tumors, whereas adult tumors mostly represented specific mature tubular cell types. RNA signals of each tumor type were remarkably uniform and specific, indicating a possible therapeutic and diagnostic utility. We demonstrated this utility with a case study of a cryptic renal tumor. Whilst not classifiable by clinical pathological work-up, mRNA signals revealed the diagnosis. Our findings provide a cellular definition of human renal tumors through an approach that is broadly applicable to human cancer.

1 **Introduction**

2

3 As cancer cells evolve from normal cells, they may retain patterns of messenger RNA

4 (mRNA) characteristic of the cell of origin. In such cases, the cancer cell transcriptome may

5 contain information that can identify the cancer cell of origin, its differentiation state, or

6 trajectory towards a cancer cell. It is therefore conceivable that tumor transcriptomes can

7 be used to identify the cells from which tumors arise and test fundamental hypotheses

8 regarding tumor's differentiation states, such as the "fetalness" of childhood tumors or the

9 dedifferentiation of adult tumors towards a fetal state.

10

11 Single cell transcriptomics allows for a direct quantitative comparison to be made between

12 single tumor and relevant normal cell transcriptomes. For example, single cell

13 transcriptomes identified that a specific subtype of proximal tubular cells are the normal cell

14 correlate of clear cell renal cell carcinoma (ccRCC) cells[1]. Such experiments can also reveal

15 more precise information about normal cells within the tumor microenvironment. However,

16 the high resource requirements of single cell transcriptomics preclude investigations of large

17 patient cohorts, which are required to study rare subtypes, test the generalizability of such

18 signals and determine associations with clinical parameters. An alternative approach is to

19 identify the presence of single cell derived mRNA signals in bulk tumor transcriptomes,

20 utilizing normal single cell transcriptomes as a reference. Smaller numbers of single cancer

21 cell experiments can then be used to validate cellular signals identified.

22

23 Tumor bulk transcriptomes for most types of human cancer have been generated in the

24 context of cancer genomics efforts of recent years, such as those conducted by the

25 *International Cancer Genome Consortium* (ICGC) and *The Cancer Genome Atlas* (TCGA)[2,3].

26 Single cell reference data, generated by efforts collectively known as the *Human Cell*

27 *Atlas*[4,5], have begun to provide quantitative transcriptional definitions of the normal cells

28 that constitute the developing and mature human kidneys[1,6–10]. By combining these bulk

29 tumor transcriptome databases with single cell reference data, we may therefore be able to

30 identify single cell signals in bulk transcriptomes across large cohorts of kidney tumors.

31

1   Here, we studied normal single cell mRNA signals in bulk kidney tumor transcriptomes

2   (n=1,258; **Fig. 1A, Table S1**) and validated our findings using targeted single cell experiments

3   (n=10, **Fig. 1A, Table S2**). There were three central aims of our analyses. Firstly, we tested

4   the fundamental presumption that childhood renal tumors exhibit fetal cell signals whilst

5   adult tumors dedifferentiate towards a fetal state. Next, we defined for each tumor type its

6   normal cell correlate which may represent its cell of origin and provide diagnostic cues.

7   Finally, we explored the tumor micro-environment across different tumor types.

8

9

1 **Results**

2

3 **An integrated single cell reference map of the kidney**

4

5 The nephron is the functional unit of the kidney and together with its associated vasculature

6 and support cells make up the majority of kidney cells. The nephron is derived from the

7 mesoderm and forms from a combination of mesenchymal cell populations that mature into

8 the epithelial cells of the nephron via mesenchymal to epithelial transition (MET)[11]. To

9 precisely define these mesenchymal populations and the populations they mature into, we

10 created a refined fetal kidney reference map combining previously generated [1,8] and newly

11 generated human fetal kidney single cell data (**Fig. 1B, S1**).

12

13 This reference revealed 4 key mesenchymal populations: mesenchymal progenitor cells

14 (MPCs), cap mesenchyme (CM), and two populations of specialized interstitial cells: smooth

15 muscle-like cells (ICa), and cortical stromal cells (ICb) (**Fig. 1B, S1A-B**) [8,11]. The cap

16 mesenchyme condenses on the ureteric bud and forms the tubular structures of the

17 nephron via mesenchymal to epithelial transition. The mesenchymal cells which do not

18 condense into cap mesenchyme and remain in the interstitial space form interstitial support

19 cells for the nephron, such as mesangial cells. The final mesenchymal population, which we

20 termed mesenchymal progenitor cells, was not present in sufficient numbers to be reported

21 in earlier single cell transcriptomic studies of the developing kidney. These MPCs are

22 enriched for early time points (**Fig. 1C**), strongly resemble mesenchymal cells in the fetal

23 adrenal (**Fig. S1D**)[12], and both populations resemble primitive mesodermal populations in

24 the post gastrulation mouse embryo (**Fig. S1D**)[13]. Developmentally, both the adrenal cortex

25 and the kidney are derived from the same mesodermal lineage.

26

27 We combined this refined map of the developing kidney with previously generated maps of

28 the mature kidney[1], the developing adrenal gland[12], and the post-gastrulation mouse[13] (**Fig.**

29 **1A**). Together these provide a complete single cell reference map of the kidney across

30 developmental time.

31

1 **Quantification of reference cellular mRNA signals in bulk transcriptomes**

2

3 Our single cell reference map of the kidney provides a cellular mRNA signal of each

4 population of cells. To measure the abundance of these reference cellular signals in bulk

5 tumor transcriptomes, we devised a method that fits raw bulk mRNA counts for the entire

6 transcriptome – not just marker genes – to a weighted linear combination of transcriptomic

7 signals derived from reference single cell data.

8

9 A number of bulk deconvolution tools exist that aim to identify the cellular composition of

10 bulk tissues using a single cell reference[14–16]. However, the aim of our analysis was not to

11 identify and quantify the number of cells present in the microenvironment, but to identify

12 the major cellular signals (or transcriptional programs) used by tumor cells. As such, we do

13 not expect any of our single cell reference populations to exactly match the tumor cells'

14 transcriptome. We therefore designed our method to identify the major transcriptional

15 signals (defined using single cell data) present in bulk transcriptomic data. We term this

16 approach "cellular signal analysis" to differentiate it from "deconvolution analysis", the

17 inference of cellular composition of bulk transcriptomes.

18

19 We applied cellular signal analysis to 766 ccRCC transcriptomes from The Cancer Genome

20 Atlas[17] to assess whether the known cellular identity of these cancer cells could be

21 identified. Our method correctly identified the signal of a specific proximal tubular cell

22 population as the predominant cell signal in ccRCC bulk cancer transcriptomes (**Fig. S2**). We

23 next applied cellular signal analysis and published deconvolution methods, MuSiC[15] and

24 BSeq-SC[14] to bulk mRNA of purified normal human B-cells, human pre-B cell leukemia, and

25 hematopoietic stem cells (**Fig. 1D**)[18,19]. For this comparison we used a reference that

26 combined single cell transcriptomes from peripheral blood cells with a negative control

27 population; proximal tubular kidney cells. BSeq-SC was unable to differentiate between

28 normal mature B cells and cancerous pre-B cell leukemia cells, giving both a mature B cell

29 label. MuSiC found an implausible renal tubular signal in both hematopoietic stem cells and

30 leukemia.

31

1   Cellular signal analysis instead explained most of the difference between the tumor and

2   reference transcriptomes through an "unexplained signal". Mathematically, this

3   "unexplained signal" represents an intercept term, included to limit the assignment of

4   spurious signals when a bulk transcriptome differs from all signals in the reference (see

5   **Methods**). Taken together, these comparisons demonstrate the need for a bespoke analysis

6   tool to perform cellular signal analysis and that the unexplained signal metric of our method

7   uniquely highlights when reference data is inappropriate.

8

9   **Childhood tumors, but not adult tumors, exhibit a fetal transcriptome**

10

11   For each tumor, we determined whether it exhibited a fetal or mature (i.e. post-natal)

12   transcriptome, to guide the choice of reference in subsequent analyses. This analysis also

13   enabled us to test two fundamental hypotheses about the differentiation state of tumors –

14   that childhood tumors represent fetal cell types and that adult cancers, especially epithelial

15   malignancies, dedifferentiate towards a fetal state.

16

17   We calculated the immaturity by fitting each bulk transcriptome to a combined reference

18   set composed of cellular signals from both mature and fetal kidney reference populations.

19   The immaturity score was the fractional contribution of the developmental signals to the

20   bulk transcriptome. Using this approach, we established a reference range of mature normal

21   kidneys (**Fig. 2A**). We demonstrated the validity of this range by scoring fetal kidney

22   transcriptomes which lay significantly outside the mature range (p=0.015, Wilcoxon rank

23   sum test).

24

25   We next calculated the same maturity score for individual tumors, which showed a clear

26   signal of "fetalness" across all types of childhood kidney tumors (**Fig. 2B-C**). Although all

27   childhood kidney tumors had a significant enrichment for developmental cellular signals,

28   pretreated Wilms tumor had a significantly lower score than other childhood kidney tumors,

29   including untreated Wilms. The comparison between treated and untreated Wilms suggests

30   that chemotherapy reduces the developmental signal in Wilms tumor, a notion we explore

31   in detail in a later section.

32

1   A significant developmental signal was absent from almost all adult tumors (**Fig. 2B**). This

2   suggests that global "dedifferentiation" to a developmental state does not occur in adult

3   kidney tumors. One obvious exception to the ubiquitous lack of a strong developmental

4   signal in adult tumors ($p<10^{-4}$, Wilcoxon rank sum test) was a cohort of lethal chromophobe

5   RCC, classified previously as metabolically divergent due to their comparatively low

6   expression of genes associated with the Krebs cycle, electron transport chain, and the AMPK

7   pathway[17].

8

9   Motivated by this observation, we tested whether other clinical markers such as somatic

10 genotype, morphology or molecularly defined subgroup were predictive of immaturity

11 score.   We found that clear cell renal cell carcinomas with two independent somatic

12 mutations in *PTEN* had a significantly higher immaturity score (**Fig. 2D**; t-test, FDR<0.01). As

13 with lethal chromophobe tumors, *PTEN* mutated ccRCCs conferred a far worse prognosis,

14 with all samples belonging to the TCGA defined m3/CCB subgroup with the worst prognostic

15 outcome of all groups[20]. Investigating further, we found an association between immaturity

16 score and the m3 transcriptional subgroup (**Fig. 2E**; t-test, FDR<0.01). No other clinical

17 covariate had a statistically significant association with immaturity score at a 1% significance

18 level (**Table S3-4**).

19

20 **Congenital mesoblastic nephroma resembles mesenchymal progenitor cells**

21

22 Congenital mesoblastic nephroma (CMN) is a renal tumor of infants that has low metastatic

23 potential. There are two morphological subtypes of CMN, classical and cellular variants[21].

24 Cell signal analysis in CMN bulk transcriptomes (n=18) revealed a uniform signal of

25 mesenchymal progenitor cells across tumors (**Fig. 3A, S3**), irrespective of morphological

26 subtype. Of note, these mesenchymal progenitor cells were characterized by expression of

27 *NTRK3* and *EGFR* genes (**Fig. 3B**), the principal oncogenes that drive CMN through activating

28 structural variants[22]. To verify that this signal was not a generic consequence of fibroblast

29 like cells, we repeated the analysis of bulk CMN transcriptomes using a developmental

30 reference combined with mature fibroblasts. This comparison revealed the same match to

31 mesenchymal progenitor cells, with a low contribution from mature fibroblasts (**Fig. 3C**).

32

1   To validate this mesenchymal stem cell signal in CMN, we subjected cells dissociated from a

2   fresh CMN tumor specimen, to single cell mRNA sequencing using the Chromium 10x

3   platform. We annotated single cells based on literature derived marker genes (**Fig. 3D, S4**)

4   and compared to single cell clusters of normal fetal kidneys using previously developed

5   quantitative approaches[1]. This comparison revealed that CMN tumor cells matched the

6   same mesenchymal progenitor cell population, validating the cell signal seen in bulk tumor

7   tissue (**Fig. 3E**).

8

9   **Wilms tumour, clear cell sarcoma of the kidney and the effect of treatment**

10

11   Wilms tumor is the most common childhood kidney cancer and is thought to arise from

12   aberrant cells of the developing nephron. Clear cell sarcoma of the kidney (CCSK) is a rare, at

13   times aggressive childhood renal cancer that is treated as a high risk Wilms tumor in clinical

14   practice[23]. We assessed the cellular signals in bulk transcriptomes from treatment-naïve

15   CCSK, high risk treatment-naïve Wilms, and intermediate risk Wilms post chemotherapy.

16   Cellular signal analysis revealed a largely uniform early nephron signal (cap mesenchyme,

17   comma shaped body, S-shaped body) in the treatment-naïve Wilms cohort (**Fig. 4A, S5**). By

18   comparison, the post-treatment cohort had a much reduced contribution from the early

19   nephron, instead of containing a mixture of tubular, early nephron and mesenchymal signals

20   with a relatively high unexplained signal fraction (**Fig. 4A, S5**). Previous work utilizing single

21   cell data from post-chemotherapy Wilms tumors identified the same lack of cap

22   mesenchyme signal identified by our analysis of bulk transcriptomes[1]. The CCSK

23   transcriptomes showed a mixture of mesenchymal and early nephron signals, with an

24   extremely high unexplained signal fraction (**Fig. 4A, S5**).

25

26   To validate the cap mesenchyme signal in treatment-naïve Wilms, we generated single cell

27   mRNA transcriptomes from one fresh sample. Annotation of this data revealed two

28   proliferating populations (**Fig. 4B, S6**). Comparison to fetal kidney showed that one of these

29   populations strongly matched the cap mesenchyme, validating its presence in treatment-

30   naïve Wilms tumor (**Fig. 4C**). The second population exhibited a strong match to

31   mesenchymal progenitor cells (**Fig. 4C**).

32

1   To further investigate the origins of CCSK we generated single nuclear transcriptomes from

2   2 archival samples and single cell transcriptomes from one fresh sample (**Fig. 4B, S7**). In

3   contrast to Wilms tumor, all CCSK tumor cells matched multiple mesenchymal and early

4   nephron populations (**Fig. 4C**). Although the matching populations were consistent with the

5   results of cell signal analysis on bulk CCSK transcriptomes (**Fig. 4A**), the match to multiple

6   reference populations at the single cell level suggests that CCSK transcriptomes represent a

7   transcriptional state that is intermediate between multiple mesenchymal populations in the

8   developing kidney. To test the possibility that the true normal cell correlate for CCSKs was

9   not in the fetal kidney, we next matched CCSK bulk transcriptomes against mature kidney,

10   fetal adrenal, developing mouse, and the pan-tissue human cell landscape[24]. In each of

11   these comparisons, the unexplained signal explained at least 50% of the CCSK bulk

12   transcriptomes, a much higher fraction than any other tumor type (**Fig. 4D**). This

13   unexplained signal fraction was comparable to the level obtained from a deliberately

14   inappropriate comparison of flow sorted B cell bulk transcriptomes compared to the non-

15   immune developing kidney (**Fig. 4D**). In aggregate, these data suggest that CCSKs represent

16   transcriptionally grossly distorted renal mesenchymal cells.

17

18   **Malignant rhabdoid tumors exhibit signals of neural crest and early mesenchyme**

19

20   Malignant rhabdoid tumor (MRT) is an aggressive, often fatal childhood cancer, that

21   typically affects the kidney but may also occur at other sites. It is considered to be the

22   extracranial counterpart of the CNS tumor, atypical teratoid/rhabdoid tumor (AT/RT). The

23   principal, usually sole, driver event in MRT and AT/RT is biallelic inactivation of *SMARCB1*. In

24   previous analyses of microRNA profiles, MRTs co-clustered with a range of tissues: neural

25   crest derived tumors, cerebellum, and synovial sarcoma[25].

26

27   Assessing fetal renal single cell signals in 65 MRT bulk transcriptomes yielded a

28   mesenchymal progenitor cell signal (**Fig. 5A, S8**). However, the nephron and unexplained

29   signal fractions were also high, indicating that tumor cells only moderately resemble this

30   reference population. To investigate further, we studied MRT single cell transcriptomes,

31   derived from an MRT expanded by a primary organoid culture (see **Methods**), from nuclear

32   mRNA sequencing, and from fresh tissue MRT cells (**Fig. 5B, S9**). Comparison to our fetal

1   kidney reference revealed that MRT cell transcriptomes did not show any consistent match

2   (**Fig. 5C**). This may indicate that the mesenchymal progenitor cell signal obtained in bulk

3   represents a signal of the broad embryological lineage of the tumor, rather than a cell type.

4

5   We therefore compared MRT cells against published reference cell populations of

6   gastrulation embryos generated from mice[13], a developmental stage that is not accessible to

7   study in humans. Although there were differences between and within samples, all

8   produced a match to neural crest and/or early mesodermal/mesenchymal populations (**Fig.**

9   **5C**). To validate this early mesodermal signal, we performed immunohistochemistry for the

10   presence of a protein specific to paraxial mesoderm, TWIST1. Consistent with its expression

11   in a subset of cells by single cell mRNA sequencing, occasional MRT cells exhibited *TWIST1*

12   staining, whilst no protein was detected in normal kidney (**Fig. 5D, S10**). Overall our data

13   show that MRTs do not exclusively exhibit mRNA signals of either neural crest or

14   mesenchyme cells. Instead, our findings point at a hybrid state of MRTs, representing mRNA

15   features of both, neural crest and mesenchyme, suggesting that MRTs may come from early

16   mesoderm or form along the differentiation trajectory of neural crest to mesenchyme.

17

18   **Adult tumors represent specific tubular cells**

19

20   As discussed above, our analyses confirmed a previous finding that the predominant single

21   cell signal in the most common types of adult renal cancer, clear cell RCC (ccRCC) and

22   papillary RCC (pRCC), was derived from a specific subtype of proximal tubular cells, termed

23   PT1 cell (**Fig. S2**)[1]. In addition, cell signal analysis also revealed some properties of the tumor

24   microenvironment. We found a prominent vascular endothelial signal in ccRCCs (**Fig. 6A,**

25   **S11**), but not in pRCCs. The downstream effects in RCC of inactivation of the von Hippel–

26   Lindau gene and upregulation of vascular endothelial growth factors are well documented[26].

27   The prominent difference in the endothelial signal provides a read-out of this pathway,

28   further explaining why anti-angiogenic treatments appear to be more effective in ccRCCs

29   than in pRCCs[27].

30

31   Continuing our investigation of the tumour microenvironment, we observed mast cells to be

32   over-represented in single cell data derived from pRCCs (**Fig. 6B**). Performing cellular signal

1   analysis revealed a high contribution of mast cell signal in a subset of tumors, significantly

2   enriched for type 1 pRCC tumors ($p<1e-4$, Wilcoxon rank-sum test; **Fig. 6C**). This finding was

3   further validated by single molecule fluorescence in-situ hybridization (smFISH), which

4   found a higher fraction of mast cells a type 1 pRCC sample, than type 2 or ccRCC (**Fig. 6D,**

5   **Table S5**).

6

7   Previous analyses of chromophobe cell renal cell carcinoma (ChRCC) have shown that ChRCC

8   exhibit expression profiles of collecting duct cells[28]. Controversy exists as to whether the

9   normal cell correlate of ChRCC is the type A or type B intercalated cells[29]. This is in part due

10   to ChRCC retaining expression of both canonical markers of intercalated cells, SLC4A1 and

11   SLC26A4 respectively (**Fig. S12**). Using cell signal analysis, which considers the entire

12   transcriptomes of type A and type B cells, rather than just two markers, revealed a uniform

13   type A signal across all chromophobe tumors (**Fig. 6E, S13**), bar the lethal variant of so-

14   called metabolically divergent tumors (**Fig. 2B, S13**). The proliferation and active remodeling

15   of type A cells has been demonstrated under conditions of systemic acidosis[30], lending

16   further credence to their possible status as the cell of origin for ChRCCs.

17

18   **Single cell signals provide diagnostic clues**

19

20   An overarching finding of our study was that each tumor type possesses a particular pattern

21   of cellular signals that were uniform in, and specific to, bulk transcriptomes from individual

22   tumor types (**Fig. 7A-B, S14**). Accordingly, cellular signal assessment of bulk transcriptomes

23   may provide sensitive and specific diagnostic clues. To test this proposition, we examined

24   cellular signals in the bulk transcriptome of a histologically undefinable metastatic primary

25   renal tumor from an 11 year old boy. Following resection, the tumor was examined

26   histologically locally and by international reference renal pathologists (**Fig. 7C**). A definitive

27   diagnosis could not be reached although an adult type renal cell carcinoma was favored.

28   Nevertheless, the child was treated as a Wilms-like tumor, with cytotoxic chemotherapy and

29   radiotherapy, following nephrectomy. He remains in complete remission two years

30   following diagnosis, thus retrospectively suggesting a diagnosis of a Wilms-like tumor, as

31   adult type kidney carcinomas do not respond to cytotoxic treatment.

32

1   We performed bulk mRNA sequencing on tumor specimens from this patient. Assessment of

2   mRNA signals in bulk tissue suggested that the tumor exhibited a fetal transcriptome with

3   cellular signals consistent with a Wilms-like tumor (**Fig. 7D-E**). The transcriptional diagnosis

4   of a Wilms-like tumor was further substantiated by analyses of whole genome sequences.

5   The tumor harbored classical somatic changes of Wilms, namely canonical *CTNNB1* and

6   *KRAS* hotspot mutations and uniparental disomy of 11p (**Fig. S15**). By comparison, when we

7   assessed single cell signals of an adult-type ccRCC that developed in a 15 year old

8   adolescent, we found an overall mature transcriptome. Furthermore, the tumor exhibited

9   the PT1 signal of ccRCC as well as a stark vascular endothelial typical of ccRCC (**Fig. 7F-G**).

10

11   **Discussion**

12

13   We have determined normal cell signals in the major types of human renal tumors. This has

14   enabled us to replace the approximate notion of the "fetalness" of childhood renal tumors

15   with quantitative transcriptional evidence that the entire spectrum of pediatric renal tumors

16   represent an aberrant developmental state. At the same time, our analyses question the

17   suggestion that adult, epithelial-derived kidney cancers dedifferentiate towards a fetal

18   state. Importantly, when we found transcriptional evidence of dedifferentiation in adult

19   tumors, it conferred a dismal prognosis. Furthermore, amongst childhood tumors we found

20   examples of cell signals representing differentiation trajectories, such as the neural crest to

21   mesenchyme conversion in MRT. By contrast, the different types of adult tumors resembled

22   specific renal tubular cells.

23

24   A central question that our findings raise is whether mRNA signals point to the cell of origin

25   of tumors. When the similarity between mRNA signals and specific cell types was high, as

26   found in most tumor types, this may be a plausible proposition. For example, in CMN, which

27   typically occurs within the first weeks of life, our analysis identified an early mesenchymal

28   progenitor cell population, characterized by the disease-defining oncogenes of CMN, as the

29   likely cell of origin of CMN. In some tumors, transformation may entirely distort and

30   obliterate gene expression profiles of the cell of origin. We found CCSK transcriptomes to

31   represent such an extreme modification of the transcriptome of the developing kidney

1

2   A further finding of our study was that within each category, the majority of tumors

3   exhibited remarkably uniform cellular signals. This indicates that there are overarching

4   transcriptional features, beyond individual gene markers, that unite tumor entities despite

5   underlying intra- and inter- tumor genetic heterogeneity. Therefore, cellular signals of renal

6   tumors may lend themselves as diagnostic adjuncts, as illustrated here by our ability to

7   resolve the identify of a histologically unclassifiable childhood tumor. Moreover, the cellular

8   transcriptome itself may represent a therapeutic target that transcends individual patients,

9   if we had tools available to manipulate transcription in a predictable manner. This may be a

10 particularly attractive approach for targeting transcriptional states of fetal cells retained in

11 childhood cancer that are absent from normal post-natal tissues.

12

13 Overall our findings attach specific cell labels to human renal tumors that are underpinned

14 by quantitative molecular data obtained from single cell mRNA sequences, independent of

15 the interpretation of marker genes. As reference data from single cell transcriptomes

16 expand through efforts such as the *Human Cell Atlas*, it will be feasible to annotate existing

17 large repositories of tumor bulk transcriptomes, to derive a cellular transcriptional definition

18 of human cancer.

19

20

Author Contributions

M.D.Y. and S.B. conceived of the experiment and wrote the manuscript. M.D.Y. performed analyses, aided by T.J.M., E.K., G.K., and T.H.H.C. I.D.V. and J.C.A. provided expertise on adrenal gland analysis. L.C. performed organoid experiments with F.A.V.B. T.R.W.O., N.S., D.R., N.C., L.H., R.K., A.W. provided pathological expertise. F.C., M.M.H.E., and A.S. provided clinical data. A.P., E.B.B., F.M., C.T., C.B., G.D.S., V.J.G., M.H., M.K., S.M.P., O.A.B., K.R., K.K, F.H., J.D., F.C.H., E.P., K.A., contributed to discussions and / or data. S.T., T.M., F.H., R.R.K, contributed fetal and tumor single cell data, together with K.B.M., R.A.B., X.H., A.W.C, L.M. S.B. and M.D.Y. directed the study, in conjunction with K.S. (single cell cancer work) and J.D. (organoid work).

Declaration of Interests

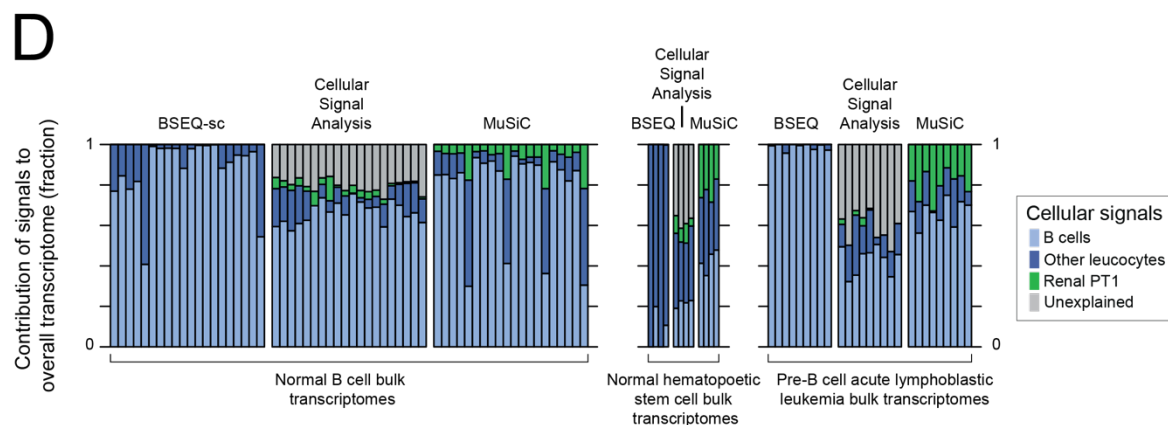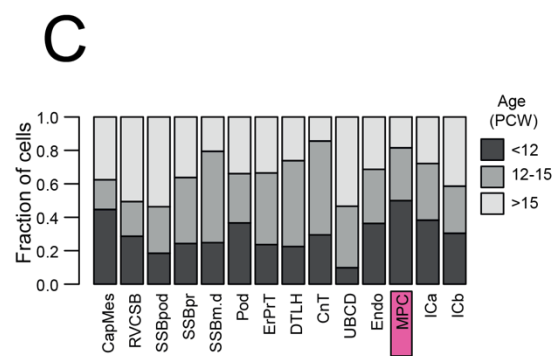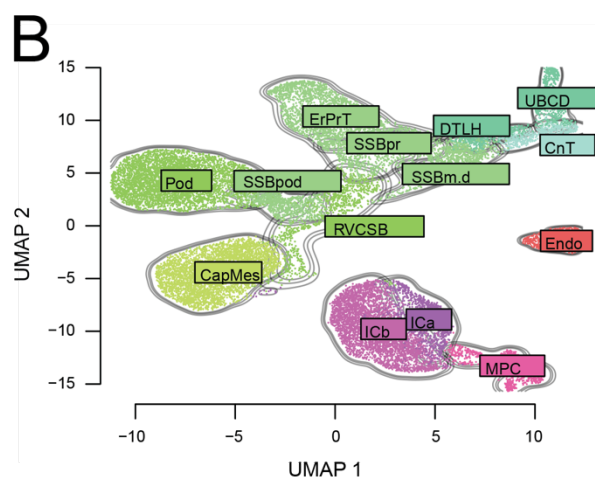The authors declare no competing interests.

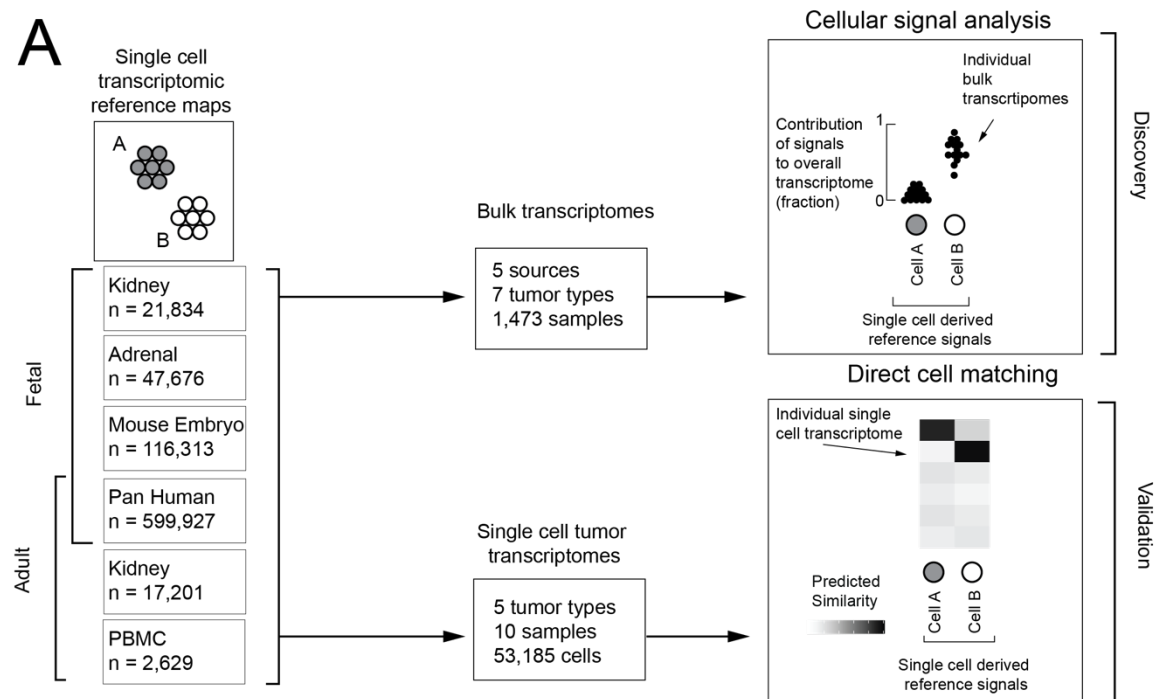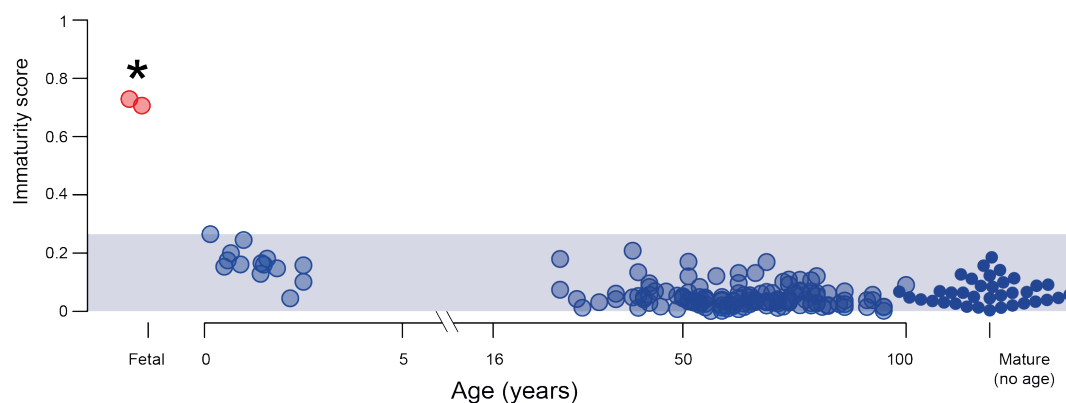1 **Figure 1**

1   **Figure 1 – Methodology overview and validation**

2

3   A. **Schematic of methodology:** Overview of the method used in this study. Single cell

4   reference signals are defined from reference single cell atlases (left). These are then

5   compared to bulk transcriptomes (top, **Table S1**) to discover the dominant normal

6   signal contribution and validated using single cell transcriptomes (bottom, **Table S2**).

7   Bulk transcriptomes are combined with reference transcriptomes to calculate the

8   relative contribution of each reference signal in explaining the bulk transcriptome

9   (see **Methods**). The strength of the signals are normalized to sum to 1, so each

10   signal (corresponding to a population of single cells in a reference single cell RNA-seq

11   dataset) is given a score between 0 and 1 for each bulk transcriptome, representing

12   the relative strength of this signal in the bulk transcriptome in question (top right).

13   For single cell transcriptomic validation, logistic regression is used to calculate a

14   similarity score for each single cell transcriptome to each reference population

15   (bottom right).

16   B. **Combined fetal kidney reference map:** UMAP representation of fetal kidney

17   reference map combining previously published and newly generated data. Contours

18   and colors indicate the labelled cell type. Abbreviations: CapMes – Cap

19   Mesenchyme, RVCSB – Renal vesicle and comma-shaped body, SSBpod – S-shaped

20   body podocyte, SSBpr – S-shaped body proximal tubules, SSBm.d – S-shaped body

21   medial and distal, Pod – Podocytes, ErPrT – Early proximal tubules, DTLH – Distal

22   tubule and loop of Henle, UBCD – Ureteric Bud and collecting duct, CnT – Connecting

23   tubules, Endo – Endothelium, ICa – Interstitial cells a (smooth muscle), ICb –

24   Interstitial cells b (stromal), MPC – Mesenchymal progenitor cells.

25   C. **Age distribution of fetal kidney populations:** Bar heights indicate the fraction of

26   cells from fetal kidney populations in **B** derived from different aged fetuses as

27   indicated by the color scale. Age is given in post conception weeks.

28   D. **Method comparison on flow sorted immune-cells and ALL:** Comparison of two

29   widely used "bulk deconvolution" methods (BSEQ-sc and MuSiC) to cellular signal

30   analysis. Each bar represents a bulk transcriptome from a population of flow sorted

31   adult B-cells (left), hematopoietic progenitors (middle), and pre-B acute

32   lymphoblastic leukemia (right). The signal contribution is calculated for each sample

1   using a reference signal set consisting of immune cells from adult peripheral blood

2   mononuclear cells and proximal tubular cells from the mature (pediatric and adult)

3   kidney (PT1 cells, included as a negative control). The relative contribution of each

4   of these reference signals (plus an "unexplained signal" component where

5   appropriate) is shown by the size and color of the stacked bars for each sample, as

6   indicated by the legend on the far right.
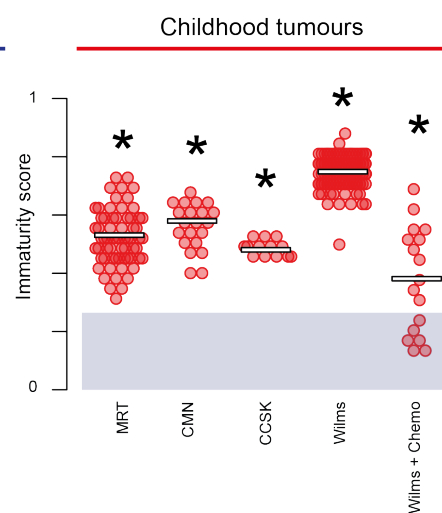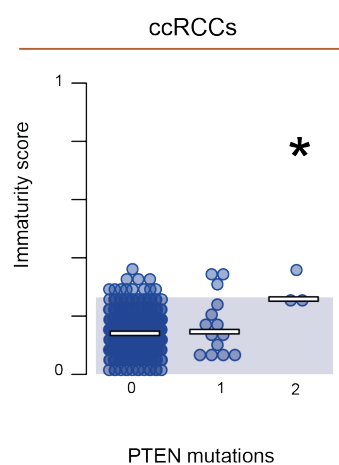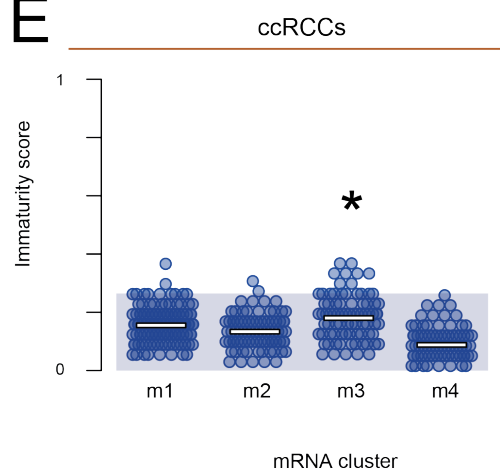
1 **Figure 2**
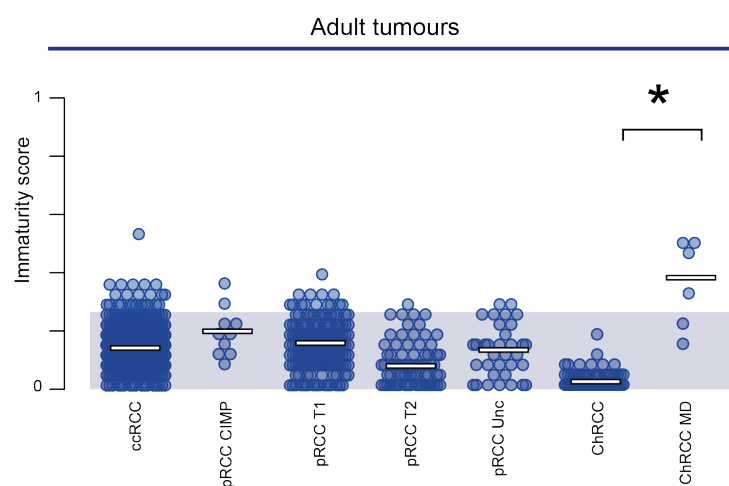
1   **Figure 2 – Immaturity score**

2

3   A. **Immaturity score for normal renal bulk transcriptomes:** For 201 bulk transcriptomes

4   derived from normal kidney biopsies across the human life span, an immaturity score

5   was calculated. This score was calculated by fitting each bulk transcriptome using a

6   combined reference set consisting of signals from all cells in the mature and fetal

7   kidney. The immaturity score is then the total normalized signal contribution from

8   fetal kidney in each bulk transcriptome (shown on the y-axis). Samples are shown by

9   age when known (x-axis), with samples for which no age information was available

10   shown on the right. The shaded blue area indicates the range of maturity scores

11   across all normal tissue post-natal transcriptomes. Bulk fetal transcriptomes are

12   shown on the left in red, with a star indicating that their maturity score is

13   significantly higher than that of the normal samples (p=0.015, Wilcoxon rank sum

14   test).

15   B. **Immaturity score for adult renal tumor transcriptomes:** The immaturity score is

16   calculated as in **A.** for 853 adult renal tumors. The normal immaturity score range is

17   shown by the blue shaded region. The metabolically divergent subtype of

18   Chromophobe renal cell carcinomas are shown separately from other chromophobe

19   RCCs. These metabolically divergent tumors have a significantly different maturity

20   score ($p<10^{-4}$, Wilcoxon rank sum test).

21   C. **Immaturity score for childhood renal tumor transcriptomes**: The immaturity score

22   is calculated as in **A.** for 287 childhood renal tumors. The normal immaturity score

23   range is shown by the blue shaded region. Each type of childhood tumor had a

24   significantly different maturity score than post-natal normal tissue kidneys ($p<10^{-4}$,

25   Wilcoxon rank sum test).

26   D. **Immaturity score for ccRCCs by PTEN mutation status:** The immaturity score for

27   clear cell renal cell carcinomas as calculated in **A**, broken down by PTEN mutation

28   status with 0 indicating wild type, 1 mono-allelic loss and 2 bi-allelic loss. The star

29   indicates that bi-allelic loss is a significant predictor of higher immaturity score

30   (p<0.01).

31   E. **Immaturity score for ccRCCs by transcriptional group:** The immaturity score for

32   clear cell renal cell carcinomas as calculated in **A**, broken down by the transcriptomic

1   subgroups identified in [20]. The star indicates that samples in m3 have a significantly
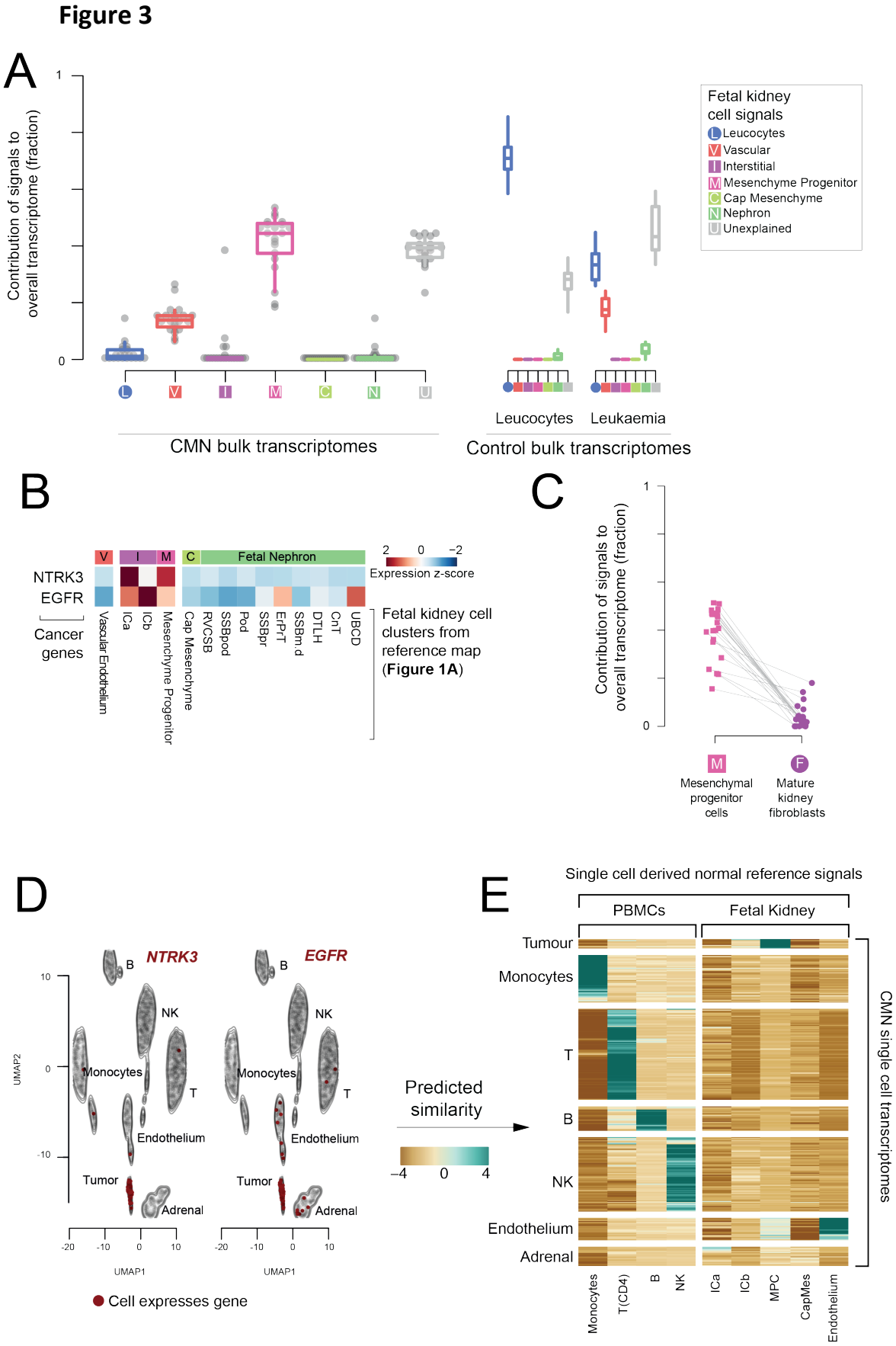
2   lower immaturity score (p<0.01).

1   **Figure 3**

2

1   **Figure 3 – Congenital mesoblastic nephromas**

2

3   A. **Composition of bulk CMNs:** The relative contribution of single cell derived signals

4   from fetal kidney in explaining the bulk transcriptomes of 18 congenital mesoblastic

5   nephromas (CMNs) along with control leucocyte and ALL populations. The relative

6   contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each

7   signal/sample combination is represented by a single point and the distribution of

8   relative signal contributions to the bulk transcriptomes are summarized with

9   boxplots. Each signal type is labelled with an abbreviation and colored as shown by

10   the legend on the right. Signals are marked with a square for fetal kidney and a

11   circle for mature. CMN samples are shown in the block on the left, while the two

12   groups of control samples are shown on the right. The samples marked "Leukocytes"

13   are bulk transcriptomes from flow sorted leukocytes, while the samples marked

14   "Leukemia" represent B-precursor acute lymphoblastic leukemia.

15   B. **Expression of CMN cancer genes in fetal kidney:** Expression of CMN driver genes

16   (rows) in reference fetal kidney single cell RNA-seq populations (columns). The data

17   has been scaled to have mean 0 and a standard deviation of 1 in each row (i.e., z-

18   transformed).

19   C. **Comparing mesenchymal progenitor cell signals to mature fibroblasts:** All 18 CMN

20   bulk transcriptomes were analyzed using a reference signal set including both fetal

21   kidney cells and the fibroblasts from mature kidney. This figure shows the

22   comparison of their inferred contribution to each transcriptome for each sample (y-

23   axis), with lines joining points representing the same sample.

24   D. **Expression of CMN marker genes:** tSNE map of single cell transcriptomes of 4,416

25   cells derived from a CMN biopsy. Cells belonging to the same cluster are indicated

26   by contours and are labelled by the cell type they represent. Cells positive for *NTRK3*

27   (left) and *EGFR* (right) are colored red. Abbreviations: B = B cell; T = T cell; DC =

28   dendritic cell; NK = NK cell; NKT = NKT cell.

29   E. **Comparison of single cell CMN to fetal kidney:** Comparison of clusters of cells for

30   which single cell transcriptomes were obtained from the CMN biopsy (rows) with

31   fetal kidney and leucocyte reference populations (columns). For each CMN

32   cluster/reference population pair a log-similarity score was calculated using logistic

1   regression (see Methods). Positive log-similarity scores represent a high probability

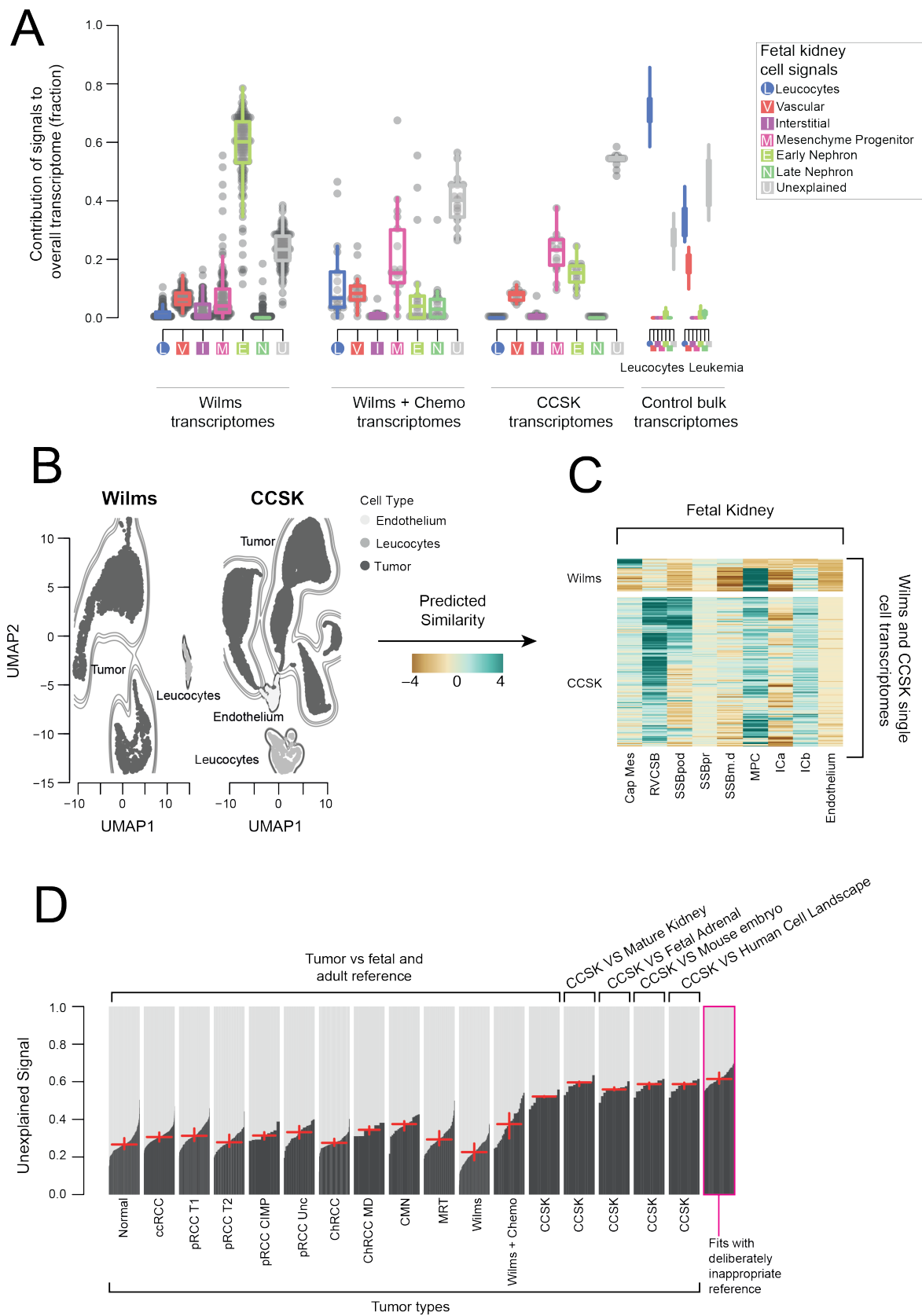2   of similarity between the reference and test cluster.

3

4

1   **Figure 4**

2

3

1   **Figure 4 – Wilms tumor and clear cell sarcoma of the kidney**

2   A. **Bulk Wilms tumor and CCSK compared to fetal kidney:** The relative contribution of

3   single cell derived signals from the fetal kidney in explaining the bulk transcriptomes

4   of 137 nephroblastomas (Wilms tumors) and 13 CCSKs along with control leucocyte

5   and ALL populations. The signal intensity assigned to each sample is shown by the y-

6   axis, normalized so that the sum of all signal contributions is 1 for each sample. The

7   relative contribution of each signal to each bulk RNA-seq sample is shown by the y-

8   axis. Each signal/sample combination is represented by a single point and the

9   distribution of relative signal contributions to the bulk transcriptomes are

10   summarized with boxplots. Each signal type is labelled with an abbreviation and

11   colored as shown by the legend on the right. Signals are marked with a square for

12   fetal kidney and circle for mature. Wilms tumor samples are shown in the block on

13   the left, while the two groups of control samples are shown on the right. The

14   samples marked "Leucocytes" are bulk transcriptomes from flow sorted leukocytes,

15   while the samples marked "Leukemia" represent B-precursor acute lymphoblastic

16   leukemia.

17   B. **UMAP of CCSK and Wilms single cell transcriptomes:** Each point represents a single

18   transcriptome from 1 Wilms tumor (left) or 3 CCSK (right) single cell transcriptome

19   samples. Shading and contours indicate the cell type, as also indicated by labels.

20   C. **Comparison of CCSK and Wilms transcriptomes to reference signals:** The similarity

21   of single cell transcriptomes of tumor populations in **B** compared to fetal kidney

22   reference signals indicated on the x-axis. Each row represents a single transcriptome

23   and the color indicates the logit similarity.

24   D. **Comparison of unexplained signal contribution to CCSKs and other tumor types:**

25   Comparison of the fraction of the bulk transcriptomes attributed to the "unexplained

26   signal" in clear cell sarcoma of the kidney samples (CCSK) and other groups of

27   samples. For each group of samples, the unexplained signal is calculated for each

28   sample individually, using the reference set of signals given at the top of the plot

29   (e.g., fetal kidney). The unexplained signal fractions are then shown by black bars,

30   sorted in increasing order, with the red horizontal line showing the median value and

31   the vertical line the range between the 25th and 75th percentiles. CCSK samples were

32   fitted using 5 different reference sets (fetal and mature kidney, mature kidney only,

1   fetal adrenal, mouse embryo, and the pan-tissue human cell landscape). The final

2   group on the right, represents samples fitted using inappropriate references. This

3   population serves as a calibration of the expected level of unexplained signal when

4   the bulk transcriptome is not explained by any of the provided reference signals.

5

1   **Figure 5**

2
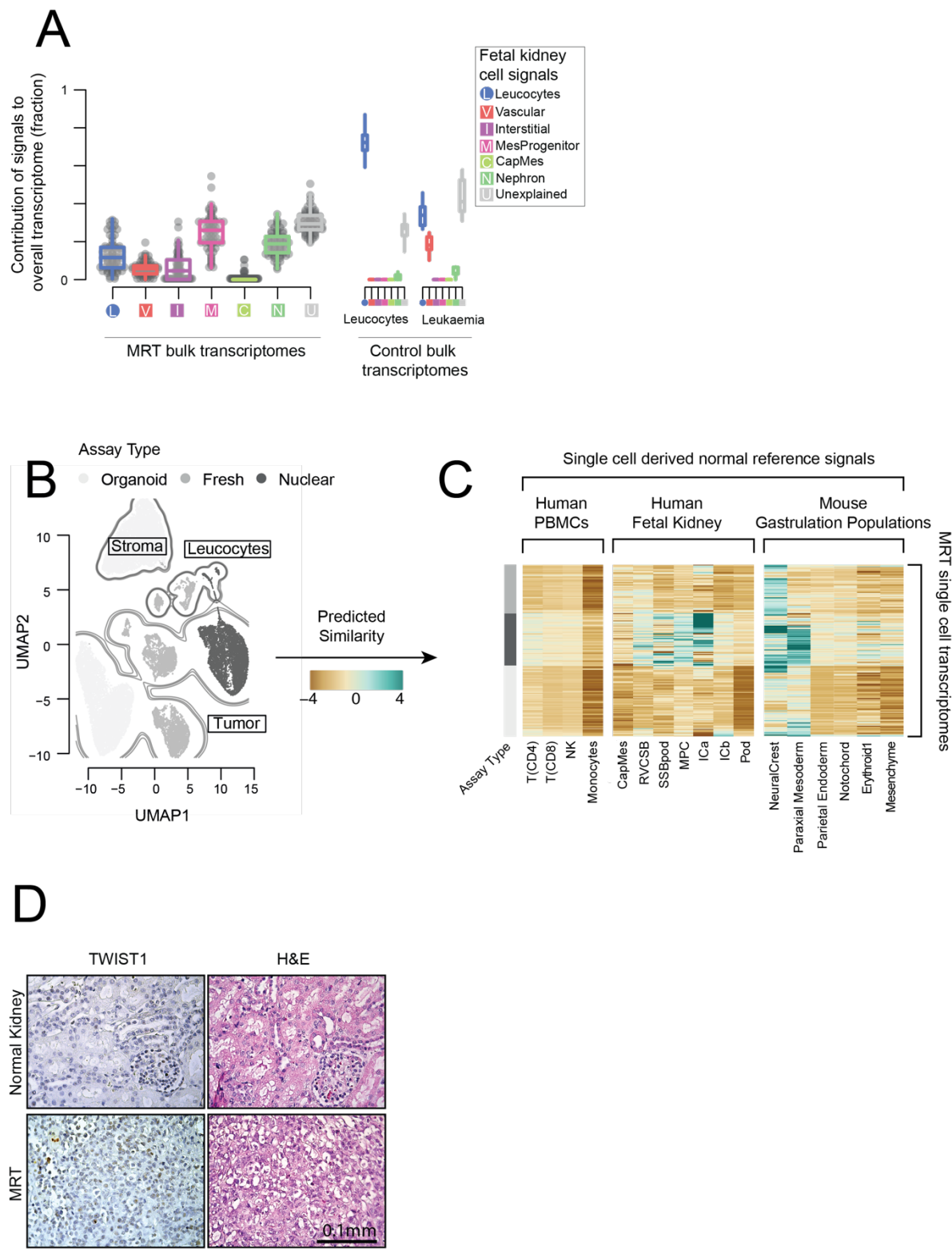
1   **Figure 5 – Malignant rhabdoid tumors**

2   A. **Bulk MRTs compared to fetal kidney:** The relative contribution of single cell derived

3   signals from fetal kidney in explaining the bulk transcriptomes of 65 malignant

4   rhabdoid tumors (MRTs) along with control leukocyte and ALL populations. The

5   relative contribution of each signal to each bulk RNA-seq sample is shown by the y-

6   axis. Each signal/sample combination is represented by a single point and the

7   distribution of relative signal contributions to the bulk transcriptomes are

8   summarized with boxplots. Each signal type is labelled with an abbreviation and

9   colored as shown by the legend on the right. Signals are marked with a square for

10   fetal kidney and circle for mature. MRT samples are shown in the block on the left,

11   while the two groups of control samples are shown on the right. The samples

12   marked "Leucocytes" are bulk transcriptomes from flow sorted leucocytes, while the

13   samples marked "Leukemia" represent B-precursor acute lymphoblastic leukemia.

14   B. **UMAP of single cell MRT transcriptomes:** Each dot represents a single transcriptome

15   from either tumor/tubular derived organoid cells (white), fresh tissue MRTs cells

16   (grey) or archival MRT nuclei (black). Contours indicate tumor cells, stroma, and

17   leucocytes as labelled.

18   C. **Log similarity of single cell MRT cells to fetal kidney and developing mouse:**

19   Comparison of the transcriptomes in **B** to cellular signals defined from single cell

20   reference transcriptomes. The reference population is indicated on the x-axis and

21   the grey bar on the left indicates the technology each cell was derived from. Each

22   row corresponds to a single transcriptome from **B**. The color scheme encodes the

23   logit similarity score for each cell against each reference population (see **Methods**).

24   D. **Immunohistochemistry of *TWIST1* in MRT and normal kidney:** Staining of a region

25   of normal kidney and MRT tissue for *TWIST1*. The MRT image shows a part of the

26   tissue selected for being *TWIST1* positive, there were large sections of tumor tissue

27   that were also *TWIST1* negative. All normal kidney tissue was *TWIST1* negative.

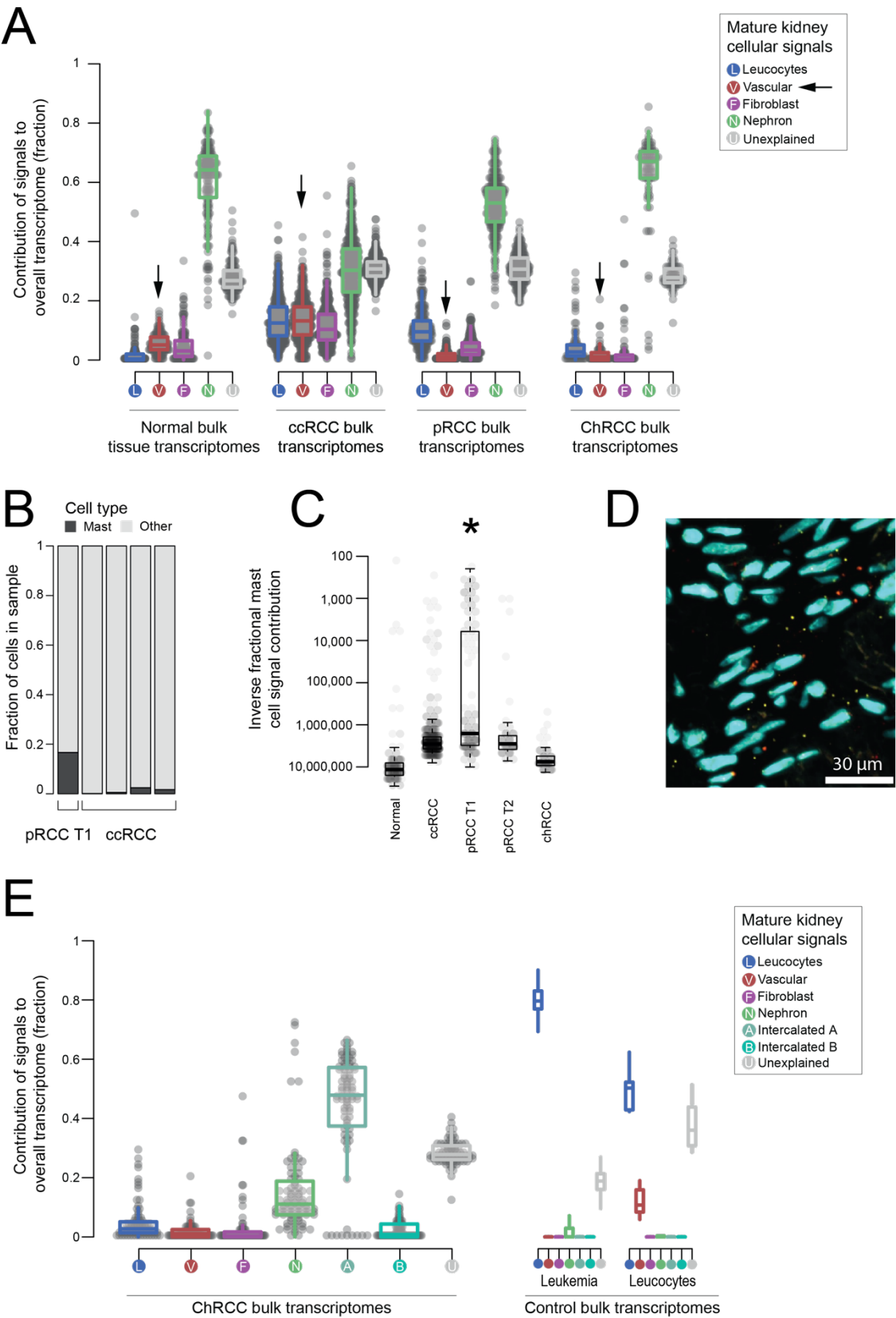28

29

30

1   **Figure 6**

2

3

1   **Figure 6 – Adult kidney tumors**

2   A. **Bulk renal cell carcinomas compared to mature kidney:** The relative contribution of

3   single cell derived signals from mature kidneys in explaining the bulk transcriptomes

4   of 171 normal kidney biopsies, 500 clear cell renal cell carcinomas (ccRCC), 274

5   papillary renal cell carcinomas (pRCC), and 81 chromophobe renal cell carcinomas

6   (ChRCC). The relative contribution of each signal to each bulk RNA-seq sample is

7   shown by the y-axis. Each signal/sample combination is represented by a single

8   point and the distributions of relative signal contributions to the bulk transcriptomes

9   are summarized with boxplots. Each signal type is labelled with an abbreviation and

10   colored as shown by the legend on the right. Signals are marked with a circle for

11   mature kidneys. The contribution to the bulk transcriptomes for all leucocyte,

12   vascular, and nephron signals are aggregated together.

13   B. **Mast cell fraction in single cell RCC samples:** Bar height indicates the fraction of cells

14   that are mast cells (black) or other (grey) in 5 renal cell carcinoma single cell

15   transcriptomic experiments (x-axis labels).

16   C. **Mast cell signals in bulk RCC transcriptomes:** Each dot represents a bulk

17   transcriptome of type indicated on the x-axis. The y-axis indicates the inverse of the

18   mast cell signal for each bulk transcriptome. Boxplots show the distribution of mast

19   cell signals across each sample type and the star indicates that mast cell signals are

20   significantly higher in pRCC T1 type tumors (Wilcoxon rank-sum test, $p < 1e-4$).

21   D. **smFISH validation:** An example section of single molecule fluorescence in-situ

22   hybridization imaging of a pRCC T1 tumor section. Nuclei are stained blue with dapi

23   and expression of the tumor marker *MET* is shown in yellow and the mast cell

24   marker *TPSB2* in red. See **Table S5** for a quantification of smFISH applied to pRCC

25   T1/T2 and ccRCC tumours.

26   E. **Bulk chromophobe renal cell carcinomas compared to mature kidney:** The relative

27   contribution of single cell derived signals from the mature kidney in explaining the

28   bulk transcriptomes of 81 chromophobe renal cell carcinomas (ChRCC), along with

29   control leucocyte and ALL populations. The relative contribution of each signal to

30   each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is

31   represented by a single point and the distribution of relative signal contributions to

32   the bulk transcriptomes are summarized with boxplots. Each signal type is labelled

1   with an abbreviation and colored as shown by the legend on the right. Signals are

2   marked with a circle for mature kidneys. ChRCC samples are shown in the blocks on

3   the left, while the two groups of control samples are shown on the right. The

4   samples marked "Leucocyte" are bulk transcriptomes from flow sorted leucocytes,

5   while the samples marked "Leukemia" represent B-precursor acute lymphoblastic

6   leukemia.
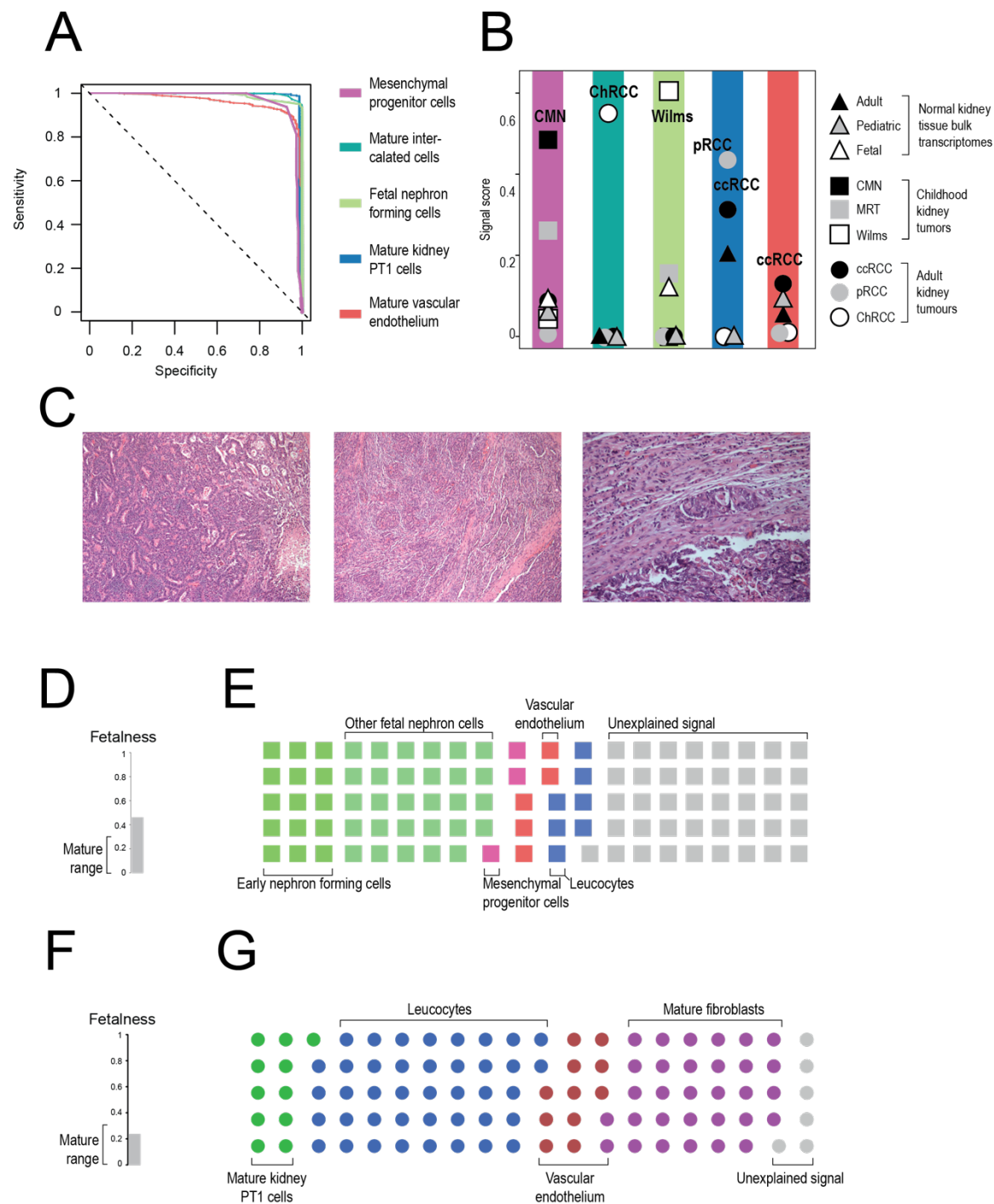
7

1   **Figure 7**

2

3

1   **Figure 7 – Clinical utility of cellular signal analysis**

2

3   A. **Sensitivity/Specificity of signals in classifying tumor types:** Curves showing the

4   sensitivity and specificity of using the scores defined by the color scheme to classify

5   tumors by type at different cut-offs. The different score and tumor type pairs are:

6   fetal interstitial cells and CMN (light blue), intercalated cells and ChRCCs (dark blue),

7   the developing nephron and Nephroblastoma (light green), PT1 and ccRCCs/pRCCs

8   (dark green), and mature vascular and ccRCCs (red).

9   B. **Median reference contribution by tumor type:** Each point represents the median

10   score for the group of samples indicated by the combination of shape and shading,

11   as explained by the legend on the right, for the score type specified by the

12   background shading. Score types are the same as in **A**.

13 C. **Histology image of unclassified childhood renal tumor:** The tumor mostly

14   compromised pleomorphic epithelioid cells that formed tubules, papillae, glands and

15   nests, as well as more solid areas with spindled cells and clefting similar to that of

16   synovial sarcoma. Patchy tumor necrosis was apparent. Some areas showed smaller,

17   more uniform cells lining narrow tubular structures, resembling adenomatous

18   perilobar nephrogenic rests. Overall, the morphology and ancillary tests were

19   inconclusive.

20 D. **Immaturity score for unclassified childhood renal tumor:** The immaturity score for

21   the unclassified childhood renal tumor, calculated as in **Fig. 2**. The range of

22   immaturity scores found in the normal mature kidney are indicated with the label on

23   the left.

24 E. **Summary of signal contribution from fetal and mature kidney to unclassified**

25   **childhood renal tumor:** The relative contribution of single cell derived signals from

26   the fetal kidney in explaining the bulk transcriptomes of the unclassified tumor from

27   **C**. The 100 squares are colored so the fraction of squares of each color matches the

28   fractional contribution of each fetal kidney signal contributes to explaining the bulk

29   transcriptome of this sample. The labels above and below indicate what each color

30   represents.

1   **F. Immaturity score for childhood renal cell carcinoma:** The immaturity score for a

2   childhood renal cell carcinoma, calculated as in **Fig. 2**. The range of immaturity

3   scores found in normal mature kidney are indicated with the label on the left.

4   **G. Summary of signal contribution from fetal and mature kidney to childhood renal**

5   **cell carcinoma:** As in **D/E** but for a transcriptome derived from renal cell carcinoma

6   fit using a mature kidney signal set.

7

1 References

2   1. Young, M. D. *et al.* Single cell transcriptomes from human kidneys reveal the cellular

3   identity of renal tumors. *Science* **361**, 594–599 (2018).

4   2. A, B., P, W. & Jc, Z. SnapShot: TCGA-Analyzed Tumors. *Cell* vol. 173

5   https://pubmed.ncbi.nlm.nih.gov/29625059/ (2018).

6   3. International Cancer Genome Consortium *et al.* International network of cancer genome

7   projects. *Nature* **464**, 993–998 (2010).

8   4. Behjati, S., Lindsay, S., Teichmann, S. A. & Haniffa, M. Mapping human development at

9   single-cell resolution. *Dev. Camb. Engl.* **145**, (2018).

10   5. A, R. *et al.* The Human Cell Atlas. *eLife* vol. 6

11   https://pubmed.ncbi.nlm.nih.gov/29206104/ (2017).

12   6. Lake, B. B. *et al.* A single-nucleus RNA-sequencing pipeline to decipher the molecular

13   anatomy and pathophysiology of human kidneys. *Nat. Commun.* **10**, 2832 (2019).

14   7. Arazi, A. *et al.* The immune cell landscape in kidneys of patients with lupus nephritis.

15   *Nat. Immunol.* **20**, 902–914 (2019).

16   8. Hochane, M. *et al.* Single-cell transcriptomics reveals gene expression dynamics of

17   human fetal kidney development. *PLoS Biol.* **17**, (2019).

18   9. Wang, P. *et al.* Dissecting the Global Dynamic Molecular Profiles of Human Fetal Kidney

19   Development by Single-Cell RNA Sequencing. *Cell Rep.* **24**, 3554-3567.e3 (2018).

20   10. Menon, R. *et al.* Single-cell analysis of progenitor cell dynamics and lineage specification

21   in the human fetal kidney. *Dev. Camb. Engl.* **145**, (2018).

22   11. Brunskill, E. W. *et al.* Atlas of gene expression in the developing kidney at microanatomic

23   resolution. *Dev. Cell* **15**, 781–791 (2008).

1   12. Kildisiute, G. *et al.* Report – Tumor to normal single cell mRNA comparisons reveal a

2   pan-neuroblastoma cancer cell. *bioRxiv* 2020.06.22.164301 (2020)

3   doi:10.1101/2020.06.22.164301.

4   13. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early

5   organogenesis. *Nature* **566**, 490–495 (2019).

6   14. M, B. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals

7   Inter- and Intra-cell Population Structure. *Cell systems* vol. 3

8   https://pubmed.ncbi.nlm.nih.gov/27667365/ (2016).

9   15. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution

10   with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).

11 16. Kang, K. *et al.* A novel computational complete deconvolution method using RNA-seq

12   data. *bioRxiv* 496596 (2018) doi:10.1101/496596.

13 17. Ricketts, C. J. *et al.* The Cancer Genome Atlas Comprehensive Molecular

14   Characterization of Renal Cell Carcinoma. *Cell Rep.* **23**, 313-326.e5 (2018).

15 18. Medeiros, B. C. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N.*

16   *Engl. J. Med.* **360**, 1787; author reply 1787-1788 (2009).

17 19. Monaco, G. *et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute

18   Deconvolution of Human Immune Cell Types. *Cell Rep.* **26**, 1627-1640.e7 (2019).

19 20. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of

20   clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).

21 21. Vujanć, G. M. & Charles, A. K. Renal tumours of childhood: an update. *Pathology (Phila.)*

22   **40**, 217–227 (2008).

23 22. Wegert, J. *et al.* Recurrent intragenic rearrangements of EGFR and BRAF in soft tissue

24   tumors of infants. *Nat. Commun.* **9**, (2018).

23. Dome, J. S. *et al.* Advances in Wilms Tumor Treatment and Biology: Progress Through International Collaboration. *J. Clin. Oncol.* **33**, 2999–3007 (2015).

24. X, H. *et al.* Construction of a human cell landscape at single-cell level. *Nature* vol. 581 https://pubmed.ncbi.nlm.nih.gov/32214235/ (2020).

25. Chun, H.-J. E. *et al.* Genome-Wide Profiles of Extra-cranial Malignant Rhabdoid Tumors Reveal Heterogeneity and Dysregulated Developmental Pathways. *Cancer Cell* **29**, 394–406 (2016).

26. O, I., Ap, L., C, J., Wg, K. & Ma, G. Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. *Proceedings of the National Academy of Sciences of the United States of America* vol. 93 https://pubmed.ncbi.nlm.nih.gov/8855223/ (1996).

27. Nm, T. *et al.* A phase 2 trial of sunitinib in patients with advanced non-clear cell renal cell carcinoma. *European urology* vol. 62 https://pubmed.ncbi.nlm.nih.gov/22771265/ (2012).

28. Davis, C. F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).

29. Lindgren, D. *et al.* Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. *Cell Rep.* **20**, 1476–1489 (2017).

30. Welsh-Bacic, D., Nowik, M., Kaissling, B. & Wagner, C. A. Proliferation of acid-secretory cells in the kidney during adaptive remodelling of the collecting duct. *PloS One* **6**, e25240 (2011).